Office of Naval Research
Contract N00014-67-A-0298-0006 NR-372-012
National Science Foundation Grant GK-31511

ON THE MINIMAX PRINCIPLE AND ZERO SUM STOCHASTIC DIFFERENTIAL GAMES

By
Yu-Chi Ho

April 1972

Technical Report No. 630

Division of Engineering and Applied Physics
Harvard University • Cambridge, Massachusetts

32

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) Division of Engineering and Applied Physics Harvard University Cambridge, Mass. 02138	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3 REPORT TITLE

ON THE MINIMAX PRINCIPLE AND ZERO SUM STOCHASTIC DIFFERENTIAL GAMES

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)
Interim technical report

5 AUTHOR(S) (First name, middle initial, last name)

Y. C. Ho

6 REPORT DATE	7a. TOTAL NO OF PAGES	7b. NO OF REFS
April 1972	32	7

8a. CONTRACT OR GRANT NO N00014-67-A-0298-0006 b. PROJECT NO c. d	9a. ORIGINATOR'S REPORT NUMBER(S) 630 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10 DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted by the U. S. Government.

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY Office of Naval Research

13 ABSTRACT

The problem of prior and delayed commitment in zero sum stochastic differential games is discussed. A new formulation and solution based on the delayed-commitment model is derived and its significant implications to stochastic game and control are considered.

DD FORM 1473 (PAGE 1)
S/N 0102-014-6700

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
prior commitment						
delayed commitment						
zero sum stochastic games						
stochastic control						
minimax						

Office of Naval Research

Contract N00014-67-A-0298-0006

NR-372-012

National Science Foundation Grant GK-31511

ON THE MINIMAX PRINCIPLE AND ZERO SUM STOCHASTIC DIFFERENTIAL GAMES

By

Yu-Chi Ho

Technical Report No. 630

April 1972

The research reported in this document was made possible through support extended the Division of Engineering and Applied Physics, Harvard University by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research and the U. S. Office of Naval Research under the Joint Services Electronics Program by Contracts N00014-67-A-0298-0006, 0005, and 0008 and by the National Science Foundation under Grant GK-31511.

Division of Engineering and Applied Physics

Harvard University · Cambridge, Massachusetts

ON THE MINIMAX PRINCIPLE AND ZERO SUM STOCHASTIC DIFFERENTIAL GAMES

By

Yu-Chi Ho

Division of Engineering and Applied Physics

Harvard University · Cambridge, Massachusetts

ABSTRACT

The problem of prior and delayed commitment in zero sum stochastic differential games is discussed. A new formulation and solution based on the delayed-commitment model is derived and its significant implications to stochastic game and control are considered.

1. Introduction

One of the fundamental tenets of game theory is the Normalization Principle of Von Neumann which roughly says that given an extensive game one can always reduce it to an equivalent game in normal form involving only strategies and payoffs and where all dynamic and informational aspects of the original problem have been suppressed in the form of strategies by considering all the possible actions of all the players under all possible circumstances. As a conceptual simplification this device is extremely useful. In fact it is so useful that one can argue that it has disproportionately influenced the development of game theory in the past two decades with the result that very little work has been done on the extensive form of games. Recently, Aumann and Maschler [1] reexamined the normalization principle and pointed out persuasively via a simple counter example of its inappropriateness under certain conditions. Their results have immediate and serious consequences in stochastic control and differential game problems since both are special cases of general extensive games. In this paper we shall:

(i) present a counter example in the same spirit as that of [1] but within the framework of a zero sum stochastic two person difference game. This example will point out the restricted circumstances under which earlier results on minimax strategies can be considered secure.

(ii) point out that (i) is actually a blessing in disguise and that from our new viewpoint we can actually solve the minimax problem for two person zero sum Linear-Quadratic-Gaussian stochastic differential (difference) games much more effectively than before. Finite dimensional

minimax solution that is eminently computable will be presented.

(iii) Show that the structure of the well known optimal stochastic control law (Kalman-Bucy filter in cascade with a zero memory linear map) for LQG problem is in fact "optimal"* under circumstances which are neither gaussian nor linear. This explains in part the incredible robustness of the LQG result in practical application and points the way to efficient solution of more general stochastic control problems.

2. The Example

The notation we shall use in this section are as follows: we write $\tilde{x}$ to denote the fact that we are considering it as a random variable, while the plain x indicates a particular sample of $\tilde{x}$; $\bar{x}$ then denotes the expected value of $\tilde{x}$; in particular, $\bar{x}'$ stands for the unconditional (prior) expectation of x and $\bar{x}''$, the conditional (posterior upon information obtained as the game evolved) expectation.

Consider the scalar two stage dynamic systems

$$\tilde{x}_3 = \tilde{x}_2 + v = (\tilde{x}_1 + u) + v \qquad \tilde{x}_1 \equiv \tilde{x} \sim N(0,\sigma) \tag{1}$$

where u and v are the controls of players I and II respectively. We have the performance criterion

$$J' = \tfrac{1}{2}E\{(\tilde{x}_3)^2 + u^2 - 2v^2\} \tag{2}$$

which I attempts to minimize and II maximize. Player I is given the measurement

$$\tilde{y} = \tilde{x} + \tilde{w} \quad , \qquad \tilde{w} \sim N(0, 1) \tag{3}$$

---

* in the sense to be explained more fully in section 6.

$\tilde{w}$, $\tilde{x}$ are independent while II receives no measurement. Both players know all the parameters and functional forms of (1)-(3). These are the common prior information.

The class of admissible strategies, Γ, for I is

$$\tilde{u} = \gamma(\tilde{z}) \qquad \gamma \in \Gamma = \text{class of all Borel measurable } \gamma : R \to R \quad (4)$$

The class of admissible strategies for II is

$$v = c \qquad c \in R \quad (5)$$

$$= \text{constant}$$

Strictly speaking, of course both γ and c depend on the common prior information such as, σ, etc. Such dependence, however, will not be explicitly shown. The expectation in (2) is taken w. r. t. the gaussian r. v. 's $\tilde{x}$, $\tilde{w}$. Using (1) and (2) we can rewrite equivalently

$$\bar{J}' = \frac{1}{2}E\{2u^2 - v^2 + 2uv + 2v\tilde{x} + 2\tilde{x}u\} \quad (2)'$$

where the term $E[\tilde{x}^2]$ is a known constant, σ, and does not enter into the game. This simple zero sum stochastic difference game can then be stated as: Find $\gamma^o \in \Gamma$, $c^o \in R$ such that (γ^o, c^o) constitutes a saddle point for $\bar{J}'$ in (2)'. This is Problem (P-1).

It is not difficult to derive that (P-1) has a saddle point in pure strategies with

$$\tilde{u}^o = \gamma^o(\tilde{z}) = -\frac{1}{2}\bar{x}'' = -\frac{1}{2}\frac{\sigma}{\sigma+1}\tilde{z} \quad (6)$$

$$v^o = c^o = 0$$

For $v = 0$

$$\min_{\gamma \in \Gamma} \bar{J}' = \min_{\gamma \in \Gamma} E_{\tilde{z}} E_{/\tilde{z}}[J] = E_{\tilde{z}} \min_{\gamma \in \Gamma} E_{/\tilde{z}}[J] \quad (7)$$

$$= E_{\tilde{z}} \min_{u} E_{/\tilde{z}}[J] \Rightarrow$$

$$u^o = -\frac{1}{2}E(\tilde{x}/z) = -\frac{1}{2}\bar{x}'' \tag{8}$$

Similarly for

$$\tilde{u}^o = v^o(\tilde{z}) = -\frac{\sigma}{2(\sigma+1)}\tilde{z} \triangleq a\tilde{z}$$

$$\underset{c}{\text{Max}}\ \bar{J}' = \underset{v}{\text{Max}}\ \frac{1}{2}E\{2a\tilde{z}^2 - v^2 + (2a\tilde{z} + 2\tilde{x})v + 2a\tilde{z}\tilde{x}\} \tag{9}$$

$$= \underset{v}{\text{Max}}\ \frac{1}{2}\{4a^2 - v^2 + 2a\} \Longrightarrow v^o = c^o = 0$$

and

$$\bar{J}'(v^o, c^o) = 2a^2 + a \tag{10}$$

The saddle point property of (v^o, c^o) is thus established. Concomitant with this saddle point property, it is often asserted or implied that if player I chooses the strategy v^o then he is guaranteed a minimax expected payoff value of (10) above. This statement has to be interpreted with considerable care as the following discussion will show. Let us consider the situation facing player I <u>after</u> he has received the information z but <u>before</u> anyone has acted. Instead of (2)', his payoff is now evaluated by

$$\bar{J}'' = \frac{1}{2}E_{/z}\{2u^2 - v^2 + 2uv + 2v\tilde{x} + 2\tilde{x}u\} \tag{11}$$

To be sure, <u>if</u> player II uses $v^o = c^o = 0$, then the optimal act for player I is still given by (8), i.e., $u^o = -\frac{1}{2}\bar{x}''$. However, <u>this action does not guarantee his security level</u> which is obtained by solving

$$\bar{J}''(u^*, v^*) = \underset{u\in R}{\text{Min}}\ \underset{v\in R}{\text{Max}}\ \bar{J}'' \tag{P-2}$$

Note that in (P-2), $\bar{x}''$ is no longer a random variable but a given number. To solve (P-2), we shall derive u^* and v^* as a saddle point pair for $\bar{J}''$.

For the purpose of solving the ZSTP game of (P-2), z can be regarded as part of the common prior information without violating the restriction of (5) on the class of admissible strategies for v. For fixed u, $\underset{v}{\text{Max}}\ \bar{J}'' \Rightarrow$

$$v^* = u + \bar{x}'' \tag{12}$$

Substituting (12) into (11) and $\underset{u}{\text{Min}}\ \bar{J}'' \Rightarrow$

$$\underset{u}{\text{Min}}\ \frac{1}{2} E_{/z} \{2u^2 - (u + \bar{x}'')^2 + 2u(u + \bar{x}'') + 2x(u + \bar{x}'') + 2\tilde{x}u\} \Rightarrow \tag{13}$$

$$u^* = -\frac{2}{3}\bar{x}'' \qquad v^* = u + \bar{x}'' = \frac{1}{3}\bar{x}'' \tag{14}$$

and

$$\bar{J}''(u^*, v^*) = -\frac{1}{6}[\bar{x}'']^2 = -\frac{1}{6}\frac{\sigma^2}{(\sigma+1)^2} z^2 \tag{15}$$

Similarly for fixed $v^* = \frac{1}{3}\bar{x}'' = \frac{1}{3}\frac{\sigma}{\sigma+1} z$, we can directly verify that $u^* = -\frac{2}{3}\bar{x}''$ is the optimal reply and yields the security level of (15). On the other hand, the strategy $\gamma^o = -\frac{1}{2}\bar{x}''$ against $v^* = u + \bar{x}''$ produces a payoff

$$\bar{J}''(\gamma^o, v^*) = \frac{1}{8}(\bar{x}'')^2 > \bar{J}''(u^*, v^*) = -\frac{1}{6}(\bar{x}'')^2 \tag{16}$$

as the case should be. The inequality of (16) is disconcerting. It says that for all possible values of z, the strategy u^* is actually a safer strategy than γ^o. Unless I has reason to believe that II has irrevocably committed himself to $v^o = c^o$ or that I can convince II that he has irrevocably committed himself to γ^o, there is no reason at all to play γ^o when u^* is safer and available. The reason for this phenomenum, as pointed out by Harsanyi [2] and Aumann and Maschler [1], is the problem of prior and delayed (posterior) commitment. Put it another way, after

the information is received we really have a nonzero sum game facing the two players with (11) the payoff for I and (2)' for II. The strategy pair (γ^o, c^o) is an equilibrium pair for I and II (in the Nash sense). However, it is well known that equilibrium strategies do not in general possess any minimax or guaranteed value properties in nonzero sum games. The above example is simply one illustration of this fact. If the game takes place at a very fast time scale such that human reactions are not practical and mechanical decision is necessary, then the prior strategy pair (γ^o, c^o) represents a reasonable solution. On the other hand, in many socio-economic multistage games, the idea of a purely mechanistic decision procedure with no human intervention and irrevocable commitment to a strategy is rather untenable when confronted with the kind of evidence in (16). In such cases, the posterior strategy u^* seems much more preferable. Of course, one may counter with the argument that since both the prior strategy and the posterior strategy for II <u>from II's viewpoint</u> are the same, $v^o = c^o = 0$, we should expect him to play it hence I should play γ^o. This reasoning is defective on two accounts:

(i) I is dependent on II's intelligence (i.e., II is clever enough to compute both the prior and posterior optimal strategies) for his payoff. But what if II is dumb but lucky to play v^*?

(ii) Suppose we endow II with the measurement

$$\tilde{y} = \tilde{x} + \tilde{\varepsilon}\ , \qquad \tilde{\varepsilon} \sim N(0,\ 1),\ \tilde{\varepsilon},\ \tilde{w},\ \tilde{x} \text{ are independent.} \tag{17}$$

then in general II will not have the same prior and posterior strategies. In fact, it can be shown that from the <u>viewpoint of player I</u>,

$$\tilde{u}^o = \gamma^o(\tilde{z}) = -\frac{\sigma(1+2\sigma)}{2(\sigma+1)^2+\sigma^2}\,\tilde{z} \qquad (18)$$

$$\tilde{v}^o = \beta^o(\tilde{y}) = \frac{\sigma(\sigma+2)}{2(\sigma+1)^2+\sigma^2}\,\tilde{y}$$

constitutes a saddle point for $\bar{\tilde{J}}'$ and

$$u^* = -\frac{2}{3}\bar{x}'' = -\frac{2\sigma}{3(\sigma+1)}\,z \qquad (19)$$

$$v^* = u + E_{/y,z}[\tilde{x}] \equiv u + \bar{x}''' = \frac{\sigma}{\sigma+1}\left[\frac{1}{2}y - \frac{1}{6}z\right]$$

is a saddle point pair for $\bar{J}''$. Note $\gamma^o(z) \neq u^*$ and $\bar{J}''(\gamma^o, v^*) > \bar{J}''(u^*, v^*)$. Furthermore, from the viewpoint of player II, he faced a payoff

$$\bar{J}''' = E_{/y}[\tilde{J}] \neq \bar{J}'' = E_{/z}[\tilde{J}] \qquad (20)$$

Since I does not have knowledge of $\bar{J}'''$, there is no compelling reason to assume that II will play γ^o unless I believes in prior commitment. In fact, the "optimal" action from II's viewpoint may just turn out to be numerically equal to v^*. In other words, I need not assume II is malicious in order to prepare for the worst.

3. Some Preliminaries to Stochastic Differential Games.

At first glance, the result of section I seems to spell disaster for practically all previous work on the stochastic (in particular Linear-Quadratic-Gaussian) differential game problem. The "minimax" or saddle point strategies that have been obtained are all of the "prior" variety. They are useful or reasonable only if we have firm belief that our opponent has made irrevocable prior commitments, before the game has begun. This severely limits their applicability not to mention the fact that in general these strategies can only be realized with infinite-

dimensional dynamic systems [3] which are hardly practical. We would like to show in below sections that our new awareness is actually a blessing in disguise and that a secure "posterior" strategy can be derived for both players that is both simple and realizable by finite dimensional linear systems.

Before we describe the problem formulation in the game situation, let us recall a few facts for the one-player linear-quadratic-gaussian stochastic control problem which we shall require later. *

Consider the finite dimensional linear stochastic dynamic system described by the Ito stochastic differential equation

$$dx = A(t)\,x\,dt + B(t)\,u\,dt + C(t)\,dw(t) \qquad x(t_0) \sim N(\hat{x}_0,\ P_0) \qquad (1)$$

$$dz = H(t)\,x\,dt + F(t)\,d\xi(t) \qquad (2)$$

where A, B, C, H, F are known nxn, nxm, nxr, pxn, pxq matrices whose elements are continuous on $[t_0,\ t_f]$ and F is of full rank with $q \geq p$ for all t. $w(t)$ and $\xi(t)$ are independent standard Wiener processes. We also consider the payoff

$$J = E(\mathcal{J}) = \frac{1}{2}\,E\left\{x(t_f)^T S_f\,x(t_f) + \int_{t_0}^{t_f} [u(t)^T R\,u(t) + x(t)^T M x(t)]\,dt\right\} \qquad (3)$$

where $S_f \geq 0$, $M(t) \geq 0$, $R(t) > 0$ are nxn, nxn, mxm, symmetric matrices whose elements are continuous on $[t_0,\ t_f]$.

First we have the following well known result.

Result 1. $x(t)$ and $z(t)$ are measurable separable gaussian random

---

* Readers well versed in control theory or engineering can skip the below technical specifications and go directly to the next section.

processes with values in R^n and R^p respectively and each having continuous sample paths with probability one [4, pp. 135-136].

Next we shall define the class of admissible control laws, Γ (strategies). Let I denote $[t_o, t_f]$; $C[t_o, t]$ the space of continuous functions on $[t_o, t]$; Z_t, the minimal σ-algebra generated by $z_t = \epsilon C[t_o, t]$ i.e.,

$$Z_t = \sigma\{Z(s),\ s\epsilon[t_o, t]\}.$$

An admissible control law is a functional $\nu: I \times C[t_o, t] \to R^m$ such that $\nu(\cdot, z_t)$ is Lebesque measurable for each $z_t \in C[t_o, t]$ and $\nu(t, \cdot)$ is Z_t-measurable for all $t \in I$. Essentially this means that the control u at t can only depend on the past and present values of the measurement history z_t. With the above set up there follows the next two well known results.

Result 2. (Kalman-Bucy Filtering) [5] The conditional mean of x(t) on Z_t, $\hat{x}(t) \triangleq E(x(t)/Z_t)$ is given by

$$d\hat{x} = (A(t)\hat{x} + B(t)u)\,dt + P(t)H^T(FF^T)^{-1}(dz - H(t)\hat{x}\,dt \qquad (4)$$

$$\hat{x}(t_o) = \hat{x}_o$$

where P(t) satisfies the DE

$$\dot{P} = AP + PA^T + CC^T - PH^T(FF^T)^{-1}HP \quad ; \qquad P(t_o) = P_o \qquad (5)$$

Corollary [5, pp. 70-72] If in addition (A, H) constitutes an observable pair, i.e.,

$$\int_t^{t_f} \xi^T(t, t_f)H^T(FF^T)^{-1}H\xi(t, t_f)\,dt > 0 \qquad \forall\, t < t_f \qquad (6)$$

where $\xi(t, \tau)$ is the fundamental matrix associated with A(t) then P(t) exists and is bounded for all $t > t_o$.

Result 3. (The Separation Principle) The optimal control law $\gamma \in \Gamma$ which minimizes (3) [5, pp. 100-101] subject to (1) and (2) is given by

$$u(t) = \gamma(t, z_t) = - R^{-1} B^T S(t) \hat{x}(t) \tag{7}$$

where

$$\dot{S} = - A^T S - SA - M + SBR^{-1} B^T S \; ; \quad S(t_f) = S_f \tag{8}$$

Corollary [5, pp. 98-99] If in addition (A, B) constitutes a controllable pair, i.e.

$$\int_{t_o}^{t} \Phi(t, \tau) B R^{-1} B^T \Phi^T(t, \tau) d\tau > 0 \qquad \forall t > t_o \tag{9}$$

then S(t) exists and is bounded for all $t < t_f$.

Operationally, what these results say is that the optimal control law can be realized by linear combinations (Eq. (7)) of the state ($\hat{x}(t)$) of a linear finite dimensional dynamic systems (Eq. (4)) which has as its input z(t). This is one of the most successful and widely used results in control theory.

In the next section we shall be using results 2 and 3 extensively. In order to avoid cumbersome notation, we shall display these two results graphically to highlight their significance. This is done in Figure 1. The optimal controller for the linear dynamics system (block ①) is another linear dynamic system of the same dimension (block ②) followed by a static linear map (block ③). Dotted lines indicate major parameter inputs to the controller which are pre-computed via Eqs. (5) and (8). In the sequel, we shall only utilize results 2 and 3 in the form of Figure 1 and avoid spelling out the various detail parameter matrices associated with each block.

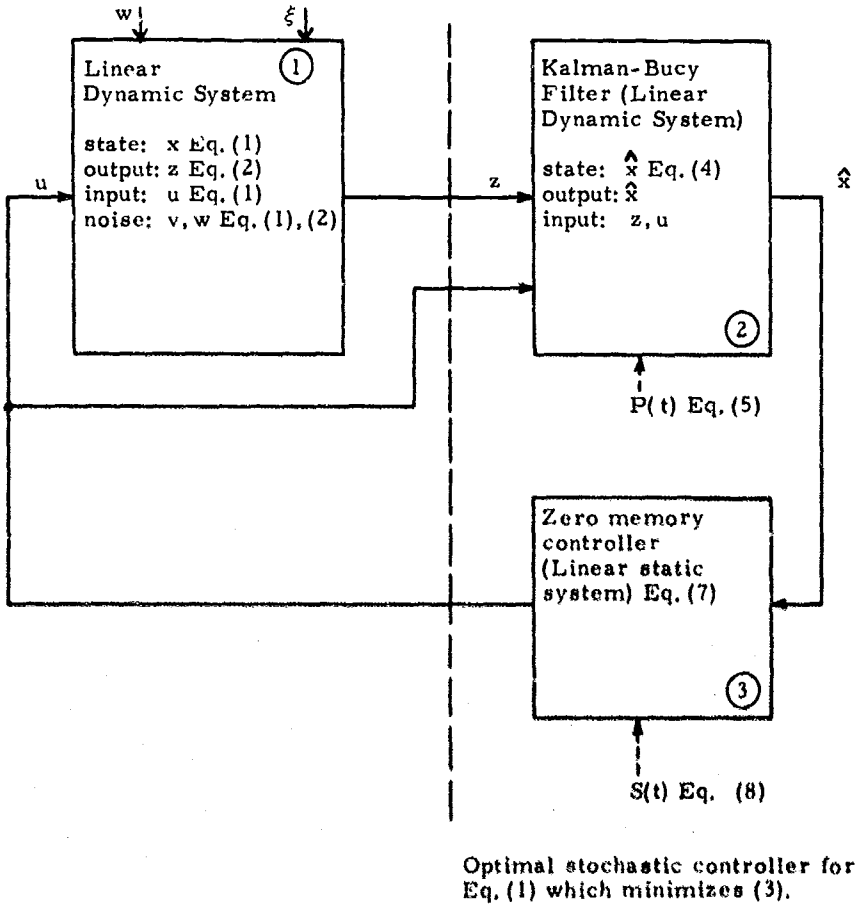

Figure 1. Graphical Representation of Results 2 and 3

4. A New Formulation and Solution of the Linear-Quadratic-Gaussian Stochastic Differential Games.

In the LQG games, instead of Eq. (3.1) we have

$$dx = (A(t)x + B(t)u + D(t)v)\,dt + C\,dw(t) \qquad (1)$$

where D is a nxs matrix similarly defined for the control input, v, player II. I and II are endowed with measurements

$$dz = Hx\,dt + Fd\beta_0(t) \tag{2a}$$

$$dy = Gx\,dt + Kd\varepsilon(t) \tag{2b}$$

respectively

where (2b) is similarly defined as (3.2) with $\varepsilon(t)$ an independent Wiener process, K is kxi ($i \geq k$) and of full rank.

The payoff is similarly defined with

$$\bar{J}' = \frac{1}{2} E\{x^T(t_f)\, S_f x(t_f) + \int_{t_o}^{t_f} [u^T R u - v^T Q v + x^T M x]dt\} \tag{3}$$

where $Q > 0$ for all $t \in I$ and the addition of $-v^T Qv$ term is due to the fact that v is maximizing. The strategy class, Γ_u, for u is same as before and Γ_v is similarly defined for v, i.e., $\beta(t, \cdot)$ is Y_t-measurable for all t and $\beta(\cdot, y_t)$ is Lebesque measurable for each $y_t \in C[t_o, t]$. The minimax strategy pair (γ^o, β^o) has been formally obtained earlier in [3]. They are infinite dimensional in the sense that block ② in Figure 2 for each player can only be realized by linear dynamic systems which are describable by partial rather than ordinary linear differential equations.

In terms of our discussions in section 2, (γ^o, β^o) are strategies of the prior commitment type. After the game has started, at time t and from the viewpoint of player I the payoff now becomes

$$\bar{J}'' = \frac{1}{2} E_{/z_t}\{J\} \tag{4}$$

While (γ^o, β^o) still retain their equilibrium property, they no longer are secure strategies. The question then arises as to what secure

strategy can player I adopt? Note that in (4), for fixed γ, β, $\bar{J}''$ is parameterized by the observation history $z_t \in Z_t$. For the purpose of computing the security payoff of (4), z_t is merely part of the prior information. It is reasonable to base the computation on the knowledge of z_t, i.e., we assume the admissible strategy class of β to include Z_t-measurable functions in addition to being Y_t-measurable. This amounts to saying that in calculating his control we shall assume that player II either through divine guidance or a perfect spy has access to player I's information. We submit that this is an eminently reasonable viewpoint to take for the purpose of calculating player I's security payoff. To be sure, we may endow player I with additional information pertaining to the problem, e.g. we may assume that II also knows w(t) or $\underline{\xi}(t)$. However, such assumptions are less natural.

Summarizing then, we wish to find $\gamma^* \in \Gamma_u$, $\beta^* \in \Gamma_u \times \Gamma_v$ such that

$$\bar{J}''(\gamma^*, \beta^*) = \min_{\Gamma \in \Gamma_u} \max_{\beta \in \Gamma_v \times \Gamma_v} [\bar{J}''] \qquad (5)$$

Our overall approach to the solution of (5) is this. We shall arbitrarily fix γ^* and then use the result of section 3 to solve

$$\bar{J}''(\gamma^*, \beta_{opt}) \geq \bar{J}''(\gamma^*, \beta), \quad \forall \beta \in \Gamma_v \times \Gamma_u \qquad (5)'$$

Let $\beta_{opt}(\gamma^*)$ be the optimal controller for II when I employs the fixed γ^*. Then fix $\beta_{opt}(\gamma^*)$ and use the result of section 3 again to solve

$$\bar{J}''(\gamma_{opt}, \beta_{opt}) \leq \bar{J}''(\gamma, \beta_{opt}) \quad \forall \gamma \in \Gamma_u \qquad (5)''$$

Let $\gamma_{opt}(\beta_{opt})$ be the solution. Consistency then requires us to solve the implicit equation

$$\gamma_{opt}(\beta_{opt}(\gamma^{*\prime})) = \gamma^{*} \tag{6}$$

Thus, let $\gamma^{*}(\cdot\,,\,\cdot)$ be a particular strategy adopted by player I. Let $\gamma^{*}(\cdot\,,\,\cdot)$ be realized by an n-dimensional linear dynamic system with state s, input z, and output u, i.e.,

$$ds = A^{*}sdt + B^{*}dz \quad , \qquad s(t_o) = \hat{x}_o \tag{7}$$

$$u = C^{*}s$$

Then Eqs. (1), (7), (2a) and (2b) appear as a combined linear dynamic system of dimension 2n (with states (x, s)) to player II through the measurements (2a) and (2b). The payoff (5) for fixed γ^{*} becomes

$$\underset{\beta \in \Gamma_u \times \Gamma_v}{\text{Max}} \; \frac{1}{2} E\left\{(x^T S_f x)_{t_f} + \int_t^{t_f} (x^T M x + s^T C^{*T} R C^{*} s - v^T Q v)dt / Z_t, Y_t\right\} \tag{8}$$

This is a standard LQG control problem to which results 2 and 3 of section 3 apply directly. The optimal controller $\beta_{opt}(t, z_t, y_t)$ is given as in Figure 2.

The combined linear dynamic system is indicated by the block ①' enclosed in dotted line. This plays the same role as block ① in Figure 1. The optimal controller, as in Figure 1, consists of blocks ②' and ③'. The filtering part, block ②' computes the estimate $\hat{x}$ and $\hat{s}$. It does this by reproducing s(t) and u(t) exactly since both s(t) and u(t) are Z_t-measurable (Hence $\hat{s}(t) \equiv E(s(t)/Z_t) = s(t)$, $\hat{u}(t) \equiv E(u(t)/Z_t) = u(t)$). The conditional mean $\hat{x}(t) \equiv E(x(t)/Z_t, Y_t)$ is computed via an n-dimensional linear system via result 2 (Kalman - Bucy filter ②s').

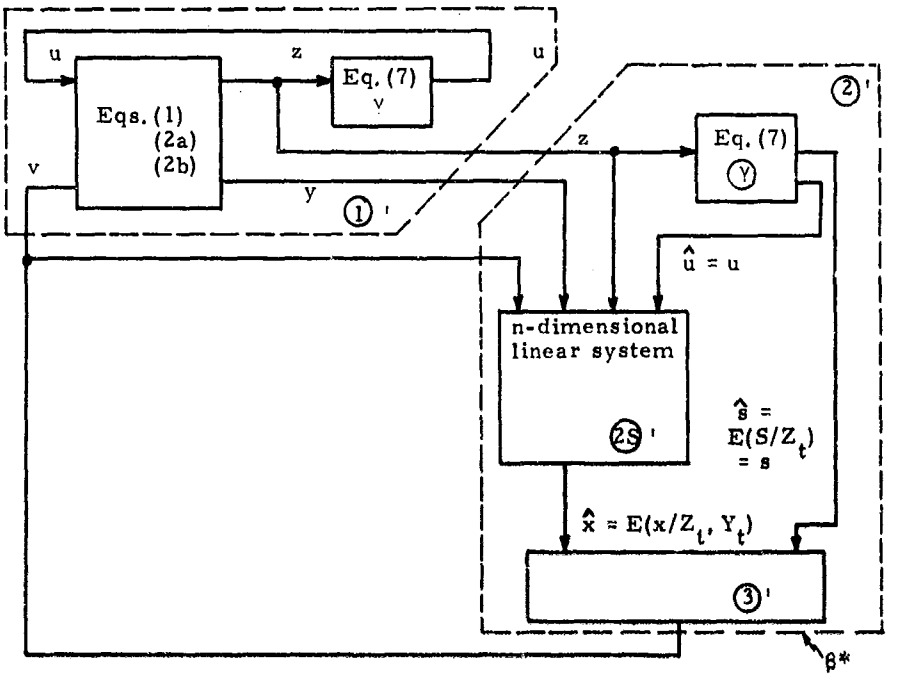

Figure 2 Optimal Controller $\beta_{opt} = \beta^*$

Block ③' is a static linear map of $\hat{x}$ and $\hat{s}$ to v similar to ③ in Figure 1, i. e.

$$v(t) = S_1(t)\hat{x}(t) + S_2(t)\hat{s}(t) \tag{9}$$

Now suppose II fixed his strategy at $\beta_{opt}(\gamma^*) \equiv \beta^*$ as determined above, we shall show that Eq. (6) precisely defines the optimal strategy for γ_{opt}. Thus γ^*, β^* constitutes a saddle point pair to (5) and

consequently solves the problem. To see this, let us consider the combined dynamic system Eqs. (1) (2) and blocks ②' ③' as appeared to I. They constitute a 3n-dimensional linear dynamic systems (with states $(x, \hat{x}, \hat{s})$): 2n from Eq. (1) and blocks ②s', and ③'; n from block Ⓥ. Furthermore, using (9) the payoff (4) becomes

$$\bar{J}' = \frac{1}{2} E\{(x^T S_f x)_{t_f} + \int_t^{t_f} (u^T R u + [x\hat{x}\hat{s}]^T [\mathcal{Q}] \begin{bmatrix} x \\ \hat{x} \\ \hat{s} \end{bmatrix}) dt / Z_t\} \tag{10}$$

where

$$\mathcal{Q} = \begin{bmatrix} M & O & O \\ O & -S_1^T Q S_1 & -S_2^T Q S_1 \\ O & -S_1^T Q S_2 & -S_2^T Q S_2 \end{bmatrix}$$

Thus I sees for fixed β^* a standard LQG problem with 3n state variables to which results 2 and 3 again apply.

We have

$$u(t) = K_1(t) E(x/Z_t) + K_2(t) E(\hat{x}/Z_t) + K_3(t) E(\hat{s}/Z_t) \tag{11}$$

However, since all outputs of block ① are Z_t-measurable by construction, they are deterministic as far as I is concerned. In fact, by definition and the requirement of Eq. (6) they are also outputs of ν^* that we are in the process of determination. Thus they need <u>not</u> be estimated or computed. The states of ①' and ②s', i.e., x and $\dot{x}$ can be estimated via result 2, i.e., we have

$$x_e \triangleq E(x/Z_t), \quad \dot{x}_e \triangleq E(\dot{x}(t)/Z_t) \triangleq E(E(x(t)/Z_t, Y_t)/Z_t) = E(x/Z_t) \triangleq x_e!$$

which are computable via a block ②'' by regarding ①', ②s', and ③'

as a new block ①". The states of ②", an n-dimensional linear dynamic system, is x_e which by construction and definition is precisely the state $\hat{s}$ of the block Ⓨ and is the conditional mean of both x and $\hat{x}$ given Z_t. Consequently from result 3, we conclude that the optimal control u can be produced using a linear combination of x_e only in a block ③" i. e. Eq. (11) becomes $u(t) = [K_1(t) + K_2(t) + K_3(t)]x_e(t)$. This is shown in Figure 3 which is simply a rearrangement of Figure 2.

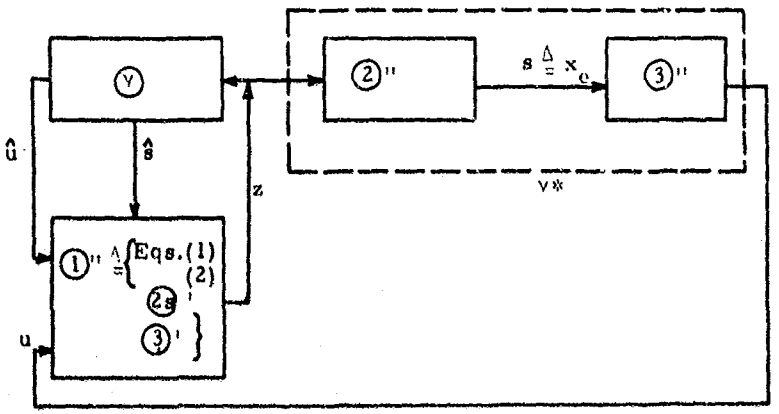

Figure 3. Optimal Controller γ*

Finally, it is worthwhile to clarify the meaning of the strategy γ* as compared to other strategies. Let (γ^o, β^o) be the minimax strategy pair determined according to [3] (the prior commitment model). At time $t = t_o$, if I has to make a commitment to a strategy for playing the rest of the

game, γ^o certainly represents a reasonable choice (similarly for β^o) since

$$\bar{J}'(\gamma^*, \beta^*) \geq \bar{J}'(\gamma^*, \beta^o) \geq \bar{J}'(\gamma^o, \beta^o) \qquad (12)^*$$

On the other hand, as soon as the game has progressed for some time, we have at $t > t_o$

$$\max_{\beta \in \Gamma_u \times \Gamma_v} \bar{J}''(\gamma^o, \beta) \geq \bar{J}''(\gamma^o, \beta^*) \geq \bar{J}''(\gamma^*, \beta^*) \qquad (13)$$

From the vantage point of I at t, γ^o becomes a rather unsafe strategy for the rest of the game compared to γ^*. To be sure, we still have $\bar{J}'(\gamma^*, \beta^*) \geq \bar{J}''(\gamma^o, \beta^o)$. But there is no compelling reason to believe that II will definitely play β^o as explained in section 2. Conceptually, at $t > t_o$, we use (γ^*, β^*) for the purpose of determining u(t) only. At $t' > t$, we have a different $\bar{J}''$ based on new information and a different minimax game to solve. A different (γ^*, β^*) will be used to determine u(t'). In general, this would require the solution of a TPZSG for each t. However, in the LQG game being discussed here, a great practical simplification occurs due to the fact that the parameters of γ^*, β^*, i.e., S_1, S_2 in Eq. (9) K_1, K_2, K_3, in Eq. (11) (see also Eqs. (5, 9) (5, 11) (5, 15) (5, 17) next section) are completely independent of z_t and y_t. Consequently, they can in fact be computed beforehand. In other words, the different (γ^*, β^*) pair I determines for each $t \geq t_o$ are in fact independent of the actual z_t. Note, however, this does not mean that we advocate I should commit himself to γ^* beforehand. Conceptually, he uses γ^* at t to

---

* Note this is different from deciding what value to use for $u(t_o)$, $v(t_o)$. In fact (γ^o, β^o) and (γ^*, β^*) will produce the same $u(t_o)$ since $Z_{t_o} = Y_{t_o}$.

compute u(t) only. He then re-solves for γ^* at each different t and uses the new (but identical) γ^* to compute the new u(t). In practice, what this means is that he must have secrecy if he decides (i.e., commits himself) to adopt the posterior strategy γ^*. He should convince his opponent that his decisions are made as the need arises and that all his options are open at all times. If no secrecy is possible and he must announce his strategy beforehand then γ^0 should be his choice.

Note that under the fictitious* saddle point condition when (γ^*, β^*) are employed, the block (Y) and γ^* are identical as well as the outputs s, $\hat{s}$ and u, $\hat{u}$. Of course, if we choose to use a different $\beta \neq \beta^*$ by say, using a $\gamma \neq \gamma^*$, in such a case $u \neq \hat{u}$ and $s \neq \hat{s}$, and $\hat{x}$, x_e can no longer be interpreted as conditional means. However, $\bar{J}''(\gamma^*, \beta^*) \geq \bar{J}''(\gamma^*, \beta)$ in this case by the derivation just given. Consequently, the minimax security level of (5) is achieved when we render β^* such that the γ block is identical to (Y*) block in Figure 3. In other words, under the conditions stated, the worst that II can do to I is to use the strategy β^*, and the best counter strategy is γ^* with $\bar{J}''(\gamma^*, \beta^*)$ the security level at time t. Of course, in real life when II does not have available both the information z(t) and y(t), I can probably expect better returns than $\bar{J}''(\gamma^*, \beta^*)$.

5. Existence Questions and a Simple Example

So far we have not addressed ourselves to the question of existence

---

*fictitious in the sense that this game is solved only for the purpose of computing I's security payoff.

of the solution which was derived in the previous section. Since the solution is obtained by solving a pair of coupled stochastic control problems (Eqs. (4. 5'), (4 5'') and 4. 6)), the existence question is directly dependent on the existence of solutions of a set of coupled Riccati equations associated with the control problems. The explicit form of these Riccati equations while straightforward to write down is rather cumbersome notationally in the general case. Nor is it possible to state simple and meaningful sufficient conditions to guarantee the existence of the solutions to these DEs. What we propose to do in this section is to carry out the derivation of the explicit solution for a very simple problem to show the various equations involved. The procedure is completely similar in the general case.

Let the scalar dynamic system and observations be

$$\dot{x} = u + v \qquad x(t_o) \sim N(\hat{x}_o,\ p_o) \tag{1}$$

$$dz = x\,dt + d\xi \tag{2}$$

$$dy = x\,dt + d\epsilon \tag{3}$$

$\xi,\ \epsilon$: are statistically independent standard wiener processes; with zero mean and variance $t - t_o$

and payoff

$$J = \frac{1}{2}x^2(t_f) + \frac{1}{2}\int_{t_o}^{t_f} (u^2 - 2v^2)\,dt \tag{4}$$

Let v^* be given by

$$ds = a\,s\,dt + b\,dz \tag{5}$$

$$u = c\,s \tag{6}$$

where a, b, and c are parameters to be determined. From I's viewpoint of a secure strategy, II maximizes at $t \geq t_o$.

$$E[\mathcal{J}/Z_t,\ Y_t] = E\{\tfrac{1}{2}x^2(t_f) + \tfrac{1}{2}\int_t^{t_f}(c^2s^2 - 2v^2)dt/Z_t, Y_t\} \tag{7}$$

subject to (1), (5), and (6). Using standard LQG results, we get for all $t \geq t_o$

$$d\hat{x} = (c\hat{s} + v)dt + p(dz + dy - 2\hat{x}\,dt) \qquad \hat{x}(t_o) = \hat{x}_o \tag{8}$$

$$d\hat{s} = a\hat{s}\,dt + b\,dz \qquad \hat{s}(t_o) = \hat{x}_o$$

where

$$\dot{p} = -2p^2 \qquad p(t_o) = p_o \tag{9}$$

and the control

$$v = \tfrac{1}{2}(S_{11}(t)\hat{x} + S_{12}(t)\hat{s}) \tag{10}$$

where

$$\dot{S}(t) \triangleq \begin{bmatrix} \dot{S}_{11} & \dot{S}_{12} \\ \dot{S}_{12} & \dot{S}_{22} \end{bmatrix} = -S\begin{bmatrix} 0 & c \\ b & a \end{bmatrix} - \begin{bmatrix} 0 & b \\ c & a \end{bmatrix}S - \begin{bmatrix} 0 & 0 \\ 0 & c^2 \end{bmatrix} - S\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}S \tag{11}$$

$$s(t_f) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Eqs. (8-11) define $\beta_{opt}(Y^*)$. Now from the viewpoint of I, β_{opt} and (1) define a 3n-dimensional linear dynamic system

$$dx = (\tfrac{1}{2}S_{11}\hat{x} + \tfrac{1}{2}S_{12}\hat{s} + u)dt \qquad x(t_o) \sim N(\hat{x}_o, P_o)$$

$$d\hat{x} = (px + (\tfrac{1}{2}S_{11} - 2p)\hat{x} + (c + \tfrac{1}{2}S_{12})\hat{s})dt + p\,dz + p\,dt \qquad \hat{x}(t_o) = \hat{x}_2 \tag{12}$$

$$d\hat{s} = a\hat{s}\,dt + b\,dt \qquad \hat{s}(t_o) = \hat{x}_o$$

with a payoff at time $t \geq t_o$

$$E\{J/Z_t\} = E\{\tfrac{1}{2}x^2(t_f) + \int_t^{t_f} [u^2 - \tfrac{1}{2}(\hat{x},\hat{s}) \begin{pmatrix} S_{11}^2 & S_{11}S_{12} \\ S_{11}S_{12} & S_{12}^2 \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{s} \end{pmatrix}]dt/Z_t\} \tag{13}$$

to be minimized. Once again using standard results, we first compute the conditional mean of x(t) and $\hat{x}(t)$, as $x_e(t)$ and $\hat{x}_e(t)$. Note that since s is Z_t-measurable we have

$$dx_e = (\tfrac{1}{2}S_{11}\hat{x} + \tfrac{1}{2}S_{12}\hat{s} + u)dt + \Sigma_{11}(t)(dz - x_e dt) \qquad x_e(t_o)=\hat{x}_o \tag{14a}$$

$$d\hat{x}_e = (px_e + (\tfrac{1}{2}S_{11} - 2p)\hat{x}_e + \tfrac{1}{2}S_{12}\hat{s} + \hat{u})dt + pdz + \Gamma_{12}(t)(dz - x_e dt) \tag{14b}$$

$$\hat{x}_e(t_o) = \hat{x}_o$$

where

$$\dot{\Gamma}(t) \triangleq \begin{bmatrix} \dot{\Sigma}_{11} & \dot{\Gamma}_{12} \\ \dot{\Gamma}_{12} & \dot{\Gamma}_{22} \end{bmatrix} = \begin{bmatrix} 0 & \tfrac{1}{2}S_{11} \\ p & \tfrac{1}{2}S_{11} - 2p \end{bmatrix} \Sigma + \Sigma \begin{bmatrix} 0 & p \\ \tfrac{1}{2}S_{11} & \tfrac{1}{2}S_{11} - 2p \end{bmatrix} \tag{15}$$

$$- \begin{bmatrix} \Sigma_{11}^2 & \Gamma_{11}\Gamma_{12} \\ \Gamma_{11}\Gamma_{12} & \Gamma_{12}^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & p^2 \end{bmatrix} \; ; \quad \Sigma(t_o) = \begin{bmatrix} P_o & 0 \\ 0 & 0 \end{bmatrix}$$

Now setting by definition $\hat{s} = s = x_e$ and noting the easily checked identity $\Sigma_{11}(t) = \Gamma_{12}(t) + P$, $\Gamma_{12}(t) = \Gamma_{22}(t)$, we can verify that

$$x_e(t) = \hat{x}_e(t) = E(x(t)/Z_t)$$

and we have finally

$$dx_e = (\tfrac{1}{2}(S_{11} + S_{12})x_e + u)dt + \Sigma_{11}(dz - x_e dt) \qquad (14)'$$

where

$$\dot{\Sigma}_{11} = \Sigma_{11}S_{11} - \Sigma_{11}^2 - pS_{11} \qquad \Sigma_{11}(t_o) = p_o \qquad (15)'$$

Furthermore,

$$u = -(K_{11}(t)x_e + K_{12}(t)\hat{x}_e + K_{13}(t)\hat{s}) \qquad (16)$$

$$= -(K_{11} + K_{12} + K_{13})x_e$$

where

$$\dot{K} = \begin{bmatrix} \dot{K}_{11} & \dot{K}_{12} & \dot{K}_{13} \\ \dot{K}_{12} & \dot{K}_{22} & \dot{K}_{23} \\ \dot{K}_{13} & \dot{K}_{23} & \dot{K}_{33} \end{bmatrix} = -KA - A^T K + \tfrac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & S_{11}^2 & S_{11}S_{12} \\ 0 & S_{11}S_{12} & S_{12}^2 \end{bmatrix} \qquad (17)$$

$$+ K\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} K \; ; \quad K(t_f) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and

$$A = \begin{bmatrix} 0 & \frac{1}{2}S_{11} & \frac{1}{2}S_{12} \\ 2p & \frac{1}{2}S_{11} - 2p & c + \frac{1}{2}S_{12} \\ b & 0 & a \end{bmatrix}$$

The consistency requirement of Eq. (3.6) now specifies

$$c \equiv -(K_{11} + K_{12} + K_{13}) \qquad (18)$$

$$b \equiv \Sigma_{11} \tag{19}$$

$$a = \frac{1}{2}(S_{11} + S_{12}) - b + c \tag{20}$$

If we substitute Eqs. (18-20) into Eqs. (9), (11), (15'), and (17), they form a set of coupled nonlinear differential equations of the Riccati type. Their solution completely specifies the secure strategy v^* via Eqs. (5), (6), (18-20). Consequently, the existence of v^* is equivalent to the existence of solutions of Eqs. (9), (11), (15') and (17).

6. Practical Implications, Open Problems, and Conclusions.

There are several implications of the results obtained that are worth further discussion.

First of all, it should be understood that the strategy we derived for u i.e., v^*, is secure only with respect to a set of assumptions which we assert to be reasonable. Roughly speaking, we allow our opponent to know everything that we may know. This seems to be as pessimistic an assumption as one would like to use. It appears paranoid to assume that the other player can have access to knowledge concerning the choices of Nature, i.e., values of $\xi(t)$, $c(t)$, $w(t)$ etc., beyond the probabilistic knowledge that are already permitted in the statement of the original problem. Our assumption is also in line with other approaches to the control of uncertain system [6, 7]. They have taken the viewpoint that such problems may be regarded as a game against an opponent (Nature) where the upper value of the game is sought. In other words, the opponent (Nature or uncertainty) makes the moves knowing everything you have known and/or have done.

In this respect, the derived solution has an additional appealing

feature. Consider the linear stochastic dynamic systems

$$dx = (Ax + Bu)dt + Cdw + vdt \tag{1}$$

$$dz = Hxdt + d\xi \tag{2}$$

where v(t) represents terms which arise due to approximations and inaccuracies in modelling of the real (probably nonlinear and nongaussian) system. Now if we consider a payoff

$$J = \frac{1}{2} E\{(x^T S_f x)_{t_f} + \int_{t_o}^{t_f} (u^T R u)dt\} \tag{3}$$

and a size-of-approximation constraints

$$E\int_{t_o}^{t_f} v^T v dt \leq \lambda \tag{4}$$

then the results in section 4 state that a "good" control law in this situation is an n-dimensional linear dynamic system followed by a zero memory linear map. This explains the almost unbelievable robustness of the structure of the well known optimal control law (Results 2 and 3) in widely diverse applications where linearity or gaussianness has been clearly violated. In other words, except for parameter values, the linear structure of v^* remains appropriate (i.e., safe) in highly nonlinear and poorly defined situations. In fact, the above discussion implies that "optimal" stochastic control of nonlinear system can now be attempted by finite dimensional optimization on the parameters of v^*. The engineering significance of this cannot be overstated.

The recognition that in the delayed commitment mode all stochastic

game in extensive form are nonzero sum raises interesting problems as well as possibilities. In this report we have only explored two solution concepts associated with NZS games, namely, Nash equilibrium and individual minimax solutions. There are many other solution concepts involving bargaining, coalitions, etc. For example, we can visualize that the two players may wish to enter into information exchange during the play of the game.

References:

1. R. Aumann and P. Mascheler, "Some Thoughts on the Minimax Principle", Research Report No. 56, Department of Mathematics, Hebrew University, Jerusalem, Israel, April 1970.

2. J. Harsanyi, "Game with Incomplete Information Played by Bayesian Players - Part II" Management Science, 14, 1968, pp. 320-334.

3. W. Willman, "Formal Solution of a Class of Stochastic Differential Games", IEEE Transactions on Automatic Control, AC-14, 1969, pp. 504-509.

4. W. Wonham, "Random Differential Equations in Control Theory", in Probabilistic Methods in Applied Mathematics, Vol. 2 (Ed. Barucha-Reid.) Academic Press 1970.

5. R. Bucy and P. Joseph, Filtering for Stochastic Processes with Applications to Guidance, Inter Science 1968.

6. H. S. Witsenhausen, "A Minimax Control Problem for Sampled Linear Systems", Transactions of IEEE Control System Society, AC-13, 1968, pp. 5-21.

7. D. P. Bertsekas and I. B. Rhodes, "On the Minimax Feedback Control of Uncertain Dynamic Systems", Proc. of 1971 IEEE Decision and Control Conference, pp. 451-455.